

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Application of:	)	
	)	
Shingo Kiuchi, et al.	)	Examiner: Not Assigned
	)	
Serial No. New Application	)	Group Art Unit No. Not Assigned
	)	
Filing Date: December 8, 2003	)	
	)	
For SPEECH RECOGNITION	)	
PERFORMANCE IMPROVEMENT	)	
METHOD AND SPEECH RECOGNITION	)	
DEVICE	)	

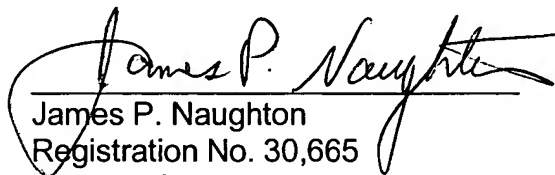
**SUBMISSION OF CERTIFIED COPY OF PRIORITY DOCUMENT**

Commissioner for Patents  
Alexandria, VA 22313-1450

Dear Sir:

Transmitted herewith is a certified copy of Japanese Patent Application No.  
2002-360336, filed on December 12, 2002.

Respectfully submitted,

  
James P. Naughton  
Registration No. 30,665  
Attorney for Applicants

BRINKS HOFER GILSON & LIONE  
P.O. BOX 10395  
CHICAGO, ILLINOIS 60610  
(312) 321-4200

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日  
Date of Application: 2 0 0 2 年 1 2 月 1 2 日

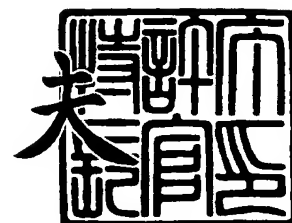
出 願 番 号  
Application Number: 特 願 2 0 0 2 - 3 6 0 3 3 6  
[ST. 10/C]: [ J P 2 0 0 2 - 3 6 0 3 3 6 ]

出 願 人  
Applicant(s): アルパイン株式会社

2 0 0 3 年 9 月 1 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



【書類名】 特許願

【整理番号】 IWP02140

【提出日】 平成14年12月12日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 5/00

【発明の名称】 音声認識装置及び音声認識性能改善方法

【請求項の数】 11

【発明者】

【住所又は居所】 東京都品川区西五反田 1 丁目 1 番 8 号 アルパイン株式会社内

【氏名】 齊藤 望

【発明者】

【住所又は居所】 東京都品川区西五反田 1 丁目 1 番 8 号 アルパイン株式会社内

【氏名】 木内 真吾

【特許出願人】

【識別番号】 000101732

【氏名又は名称】 アルパイン株式会社

【代理人】

【識別番号】 100084711

【弁理士】

【氏名又は名称】 齊藤 千幹

【電話番号】 043-271-8176

【手数料の表示】

【予納台帳番号】 015222

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】

要約書 1

【プルーフの要否】

要

【書類名】 明細書

【発明の名称】 音声認識装置及び音声認識性能改善方法

【特許請求の範囲】

【請求項 1】 音声の認識性能を改善する音声認識装置における音声認識性能改善方法において、

音声認識対象である音声データより非音声区間の始点位置が異なる複数の音声データを生成し、

前記各音声データを用いて音声認識を行い、

得られた複数の認識結果のなかで、最も多い認識結果を出力する、

ことを特徴とする音声認識性能改善方法。

【請求項 2】 前記非音声区間の始点位置を、音声区間の始点位置から所定時間前の位置まで順次移動させることにより、前記音声認識対象である音声データより非音声区間の始点位置が異なる複数の音声データを生成する、

ことを特徴とする請求項 1 記載の音声認識性能改善方法。

【請求項 3】 前記音声区間の始点位置を、前記音声認識を行う音声認識エンジンより出力する、

ことを特徴とする請求項 2 記載の音声認識性能改善方法。

【請求項 4】 入力した音声信号を所定サンプリング時間間隔で AD 変換してバッファにサンプリング順に格納し、該バッファからの読み出し位置を変えることにより非音声区間の始点位置が異なる複数の音声データを生成する、

ことを特徴とする請求項 1 又は 2 記載の音声認識性能改善方法。

【請求項 5】 非音声区間の始点位置が異なる複数の音声データのそれぞれに対応して音声認識エンジンを設け、各音声認識エンジンの認識結果のなかで、最も多い認識結果を出力する、

ことを特徴とする請求項 1 記載の音声認識性能改善方法。

【請求項 6】 音声の認識性能を改善する音声認識装置において、

音声認識対象である音声データより非音声区間の始点位置が異なる複数の音声データを生成する音声データ生成部、

前記各音声データを用いて音声認識を行う音声認識エンジン、

得られた複数の認識結果のなかで、最も多い認識結果を選択して出力する認識結果選択部、

を備えたことを特徴とする音声認識装置。

【請求項7】 前記音声データ生成部は、  
前記非音声区間の始点位置を、音声区間の始点位置から所定時間前の位置まで、順次移動させることにより、前記音声認識対象である音声データより非音声区間の始点位置が異なる複数の音声データを生成する、  
ことを特徴とする請求項6記載の音声認識装置。

【請求項8】 前記音声認識エンジンは、前記音声区間の始点位置を出力する、  
ことを特徴とする請求項7記載の音声認識装置。

【請求項9】 入力した音声信号を所定サンプリング時間間隔でAD変換するADコンバータ、  
AD変換された音声データをサンプリング順に格納する音声バッファ、  
を備え、前記音声データ生成部は、該バッファからの読み出し開始位置を変えることにより非音声区間の始点位置が異なる複数の音声データを生成する、  
ことを特徴とする請求項6又は7記載の音声認識装置。

【請求項10】 音声の認識性能を改善する音声認識装置において、  
音声認識対象である音声データより非音声区間の始点位置が異なる複数の音声データを生成する音声データ生成部、  
非音声区間の始点位置が異なる複数の音声データのそれぞれに対応して設けられ、該音声データを用いて音声認識を行う音声認識エンジン、  
各音声認識エンジンの認識結果のなかで、最も多い認識結果を選択して出力する認識結果選択部、  
を備えたことを特徴とする音声認識装置。

【請求項11】 入力した音声信号を所定サンプリング時間間隔でAD変換するADコンバータ、  
AD変換された音声データをサンプリング順に格納する音声バッファ、  
を備え、前記音声データ生成部は、該バッファからの読み出し開始位置を変える

ことにより非音声区間の始点位置が異なる複数の音声データを生成して各音声認識エンジンに入力する、

ことを特徴とする請求項10記載の音声認識装置。

#### 【発明の詳細な説明】

#### 【0001】

##### 【発明の属する技術分野】

本発明は、音声認識装置及び音声認識性能改善方法に係わり、特に、ノイズ環境下での音声認識性能を改善する音声認識装置及び音声認識性能改善方法に関する。

#### 【0002】

##### 【従来の技術】

音声によってオーディオ機器、ナビゲーションシステムなど車載機器の操作を行う、音声認識装置が実用化されている。図6はかかる音声認識装置のブロック構成図であり、音声入力用マイクロホン1は話者が入力した音声を検出して出力し、ADコンバータ2は音声信号をデジタルに変換し、操作部3は図示しないスイッチ操作により音声認識開始を指示し、音声認識エンジン4は音声認識開始が指示されたとき、入力音声进行認識する。

音声認識エンジン4の一例が特開昭59-61893号公報に開示されている。この従来技術では、単語入力音声における一連の単音節毎の特徴パターンを標準パターンと比較して音声認識し、単語辞書を参照して認識結果を意味のある単語として出力するものである。

【特許文献1】 特開昭59-61893号公報

#### 【0003】

##### 【発明が解決しようとする課題】

音声認識システムに入力する音声データにノイズが重畳されている場合、データの先頭部分である非音声区間の一部分を削除するなどして(非音声区間の長さを変更して)、音声区間の先頭位置を変えて音声認識エンジンに入力すると、認識結果が変わることがある。すなわち、同じ発声でも、その発話タイミング(音声区間の先頭位置)によって、認識結果の正誤が変わってしまう。

## 【0004】

この現象は、音声データに重畳するノイズ、例えば車室内ノイズの大きさが音声に対して十分に小さい（SN比が高い）場合は、ほとんど現れないが、車室内ノイズの大きさが音声に対して大きい（SN比が低い）場合には、顕著に現れる。このような現象が発生する理由としては、音声認識エンジン4が、非音声区間\*SIT(図7)においてバックグラウンドのノイズレベルを測定し、音声区間SITの音声データより音声認識処理する際、該ノイズレベルを用いているからである。なお、非音声区間\*SITは、スイッチで音声認識開始を指示した時刻 $t_B$ から音声区間SITの始点位置（発話タイミング） $t_{TS}$ までの区間である。

## 【0005】

このノイズデータの測定は、短時間区間の測定のため、同じ条件下のノイズでも、測定位置によって測定結果にバラツキを生じてしまう。そのため、認識結果にもバラツキが発生し、結果として正しく認識したり、誤認識したりしてしまうと考えられる。例えば、図7に示すようにノイズレベルを非音声区間\*SITの平均レベルとし、該ノイズレベルを考慮して音声区間SITにおける音声データを用いて音声認識するものとする、図では非音声区間\*SITの先頭でノイズレベルが大きいためノイズの平均レベルは非音声区間\*SITが短いほど、すなわち発話タイミング $t_{TS}$ が早いほど平均レベルが大きくなり、長いほど（発話タイミング $t_{TS}$ が遅いほど）小さくなる。以上のように、発話タイミング $t_{TS}$ により測定するノイズレベルが変化し、これにより、認識結果の正誤が変わってしまうのである。

## 【0006】

以上の現象は、ある程度のS/Nが確保されている環境であっても、発話のタイミングのせいで誤認識してしまう事態を呈し、ユーザー側からみると、認識性能が低下したことと同じであり、問題である。

特開昭59-61893号公報の技術を含め、従来技術では、専ら音声認識エンジンの認識精度を高めて認識率向上を目指しているが、限度がある。

以上から本発明の目的は、音声認識エンジンを変えずに音声認識性能を向上できるようにすることである。

## 【0007】



**【課題を解決するための手段】**

音声認識装置において、音声認識対象である音声データより非音声区間の始点位置が異なる複数の音声データを生成し、前記各音声データを用いて音声認識を行い、得られた複数の認識結果のなかで、最も多い認識結果を出力する。このようにすれば、非音声区間の始点位置をシフトするため、誤認識する音声データはあるかもしれないが、多数の音声データを認識してその数を比較すれば、正しく認識する場合が一番多くなるはずであり、従って、一番多く認識された結果を出力することにより認識性能を、認識エンジンを変えずに向上することができる。

非音声区間の始点位置が異なる複数の音声データを生成するには、前記非音声区間の始点位置を、音声区間の始点位置から所定時間前の位置まで順次シフトさせる。すなわち、入力した音声信号を所定サンプリング時間間隔でAD変換してバッファにサンプリング順に格納し、該バッファからの読み出し位置を変えることにより非音声区間の始点位置が異なる複数の音声データを生成する。

前記各音声データの音声認識処理を1台の認識エンジンが行ってもよいが時間を要する。処理時間を短縮するために非音声区間の始点位置が異なる複数の音声データのそれぞれに対応して音声認識エンジンを設け、各音声認識エンジンの認識結果のなかで、最も多い認識結果を出力する。

以上のように本発明の音声認識装置によれば、認識エンジンを変えずに音声認識性能を向上することができる。

**【0008】****【発明の実施の形態】****(A) 第1実施例**

図1は本発明の第1実施例の音声認識装置の構成図、図2は音声データ生成部の説明図である。スイッチ動作で音声認識開始が指示されると、音声入力用マイクロホン11は話者が入力した音声を検出して出力し、ADコンバータ12は入力された音声信号(図3参照)を所定サンプリング速度でAD変換し、音声バッファ13はAD変換された音声データをサンプリング順に格納する。音声データは図3に示すように非音声区間\*SITの音声信号(ノイズ)と音声区間SITの音声信号とを時系列にサンプリングして作成され、順に番号1～nが付されて図2に示すように音声バッファ13に順

番に保存される。番号が若い方は非音声区間\*SITのデータ、番号が大きい方は音声区間SITのデータである。

#### 【0009】

音声データ生成部14は、音声バッファ13からの読み出し開始位置をシフトすることにより、非音声区間の始点位置が異なる複数の音声データDT1, DT2, DT3, ……を生成して音声データ記憶部15に格納する。読み出し開始位置のシフト点は、図3に示すように音声区間SITの始点位置 $t_{ST}$ から所定時間T前の位置 $t_C$ までである。認識処理部16の音声認識エンジン17は音声区間SITの始点位置 $t_{ST}$ を識別し、該始点位置 $t_{ST}$ を用いて最終シフト位置 $t_C$ を求めて音声始端検出情報SSTとして出力する。音声始端検出情報SSTは、一つの音声データを認識処理する毎に得られるが、最初の音声データを認識処理したときの始端検出情報を使っても良いし、最初から数データ分の音声始端検出情報を平均したものを使っても良い。

#### 【0010】

図2を参照すると、音声データ生成部14のポインタ制御部21は音声認識エンジン17から入力する音声始端検出情報SST (=読み出し開始の最終シフト位置 $t_C$ )を受信する。ポインタ制御部21は読み出し位置(ポインタ)を順次シフトしてデータ読み出し部22に入力し、データ読み出し部22は指示されたポインタの示す位置からサンプリングデータを音声バッファ13から読み出して音声データ記憶部15に保存する。1つの音声データの読み出しが完了すれば、ポインタ制御部21は読み出し開始位置を1サンプリングデータ分シフトし、読み出し位置(ポインタ)を順次シフトしてデータ読み出し部22に入力し、データ読み出し部22は入力されたポインタの示す位置からサンプリングデータを音声バッファ13から読み出して音声データ記憶部15に保存する。以後、音声データの読み出しが完了する毎に読み出し位置(ポインタ)をシフトして音声データの読み出しを行い、読み出し位置(ポインタ)が最終シフト位置 $t_C$ に等しくなったとき音声データの生成処理が完了する。

#### 【0011】

以上と並行して、認識処理部16の音声認識エンジン17は、第1番目の音声データDT1を用いて音声認識処理を行うと共に、音声区間SITの始点位置 $t_{ST}$ を検出して音声始端検出情報SSTを出力する。そして認識結果(認識結果1)を認識結果保存部18

に保存する。

ついで、音声認識エンジン17は、第2番目の音声データDT2を用いて音声認識処理を行って認識結果2を認識結果保存部18に保存する。以後、同様にして全音声データDT1～DTkの認識結果1～kを認識結果保存部18に保存する。

全音声データDT1～DTkの認識が終了すれば、集計／比較部19は、得られた複数の認識結果のなかで、最も多い認識結果を最終結果として出力する。図4は集計／比較部19の構成図であり、集計部31と比較結果出力部32を有している。集計部31は比較結果毎にその数を集計する。図では、比較結果A, B, Cの数はp, q, rである。比較結果出力部32はp, q, rのうち最大値に応じた認識結果を最終認識結果として出力する。

#### 【0012】

以上、第1実施例によれば、非音声区間の始点位置をシフトするため、ノイズの影響により、たまたま誤認識する音声データはあるかもしれないが、ノイズが存在しないときには正しく認識する音声認識エンジンであれば、多数の音声データを認識してその数を比較すれば、正しく認識する場合が最も多くなる。従って、一番多く認識された結果を出力することにより認識性能を、認識エンジンを変えずに向上することができる。

#### 【0013】

##### (B) 第2実施例

図5は第2実施例の音声認識装置の構成図であり、図1の第1実施例と同一部分には同一符号を付している。第1実施例と異なる点は、

- (1) 音声データDT1～DTkを保存する音声データ記憶部15を削除した点、
  - (2) 音声データ生成部14から出力するk個の音声データに対応してk個の認識エンジン17<sub>1</sub>～17kを設けている点、
  - (3) 各認識エンジン17<sub>1</sub>～17kがk個の音声データに対して音声認識処理を行って、それぞれ認識結果A, B, C...を集計／比較部19に入力する点、
  - (4) 集計／比較部19が各音声認識エンジン17<sub>1</sub>～17kの認識結果のなかで、最も多い認識結果を最終認識結果として出力する、
- 点である。

このようにk個の音声認識エンジンを設けることにより高速の音声認識結果を得ることができ、しかも、認識性能を向上することができる。

#### 【0014】

##### 【発明の効果】

以上本発明によれば、音声認識対象である音声データより非音声区間の始点位置が異なる複数の音声データを生成し、前記各音声データを用いて音声認識を行い、得られた複数の認識結果を出力するように構成したから、認識エンジンを変えずに認識性能を向上することができる。

又、本発明によれば、複数の音声データのそれぞれに対応して音声認識エンジンを設けることにより高速の音声認識結果を得ることができ、しかも、認識性能を向上することができる。

又、本発明によれば、ある程度以上のS/N（2～3dB以上）が確保されている環境下で認識エンジンを用いたときに、発話タイミングのせいで誤認識するという現象をなくすことができる。これは、ユーザ側からみると、ノイズ環境下での認識性能が向上することと同じ効果をもつことになり本発明は有用である。

##### 【図面の簡単な説明】

##### 【図1】

本発明の第1実施例の音声認識装置の構成図である。

##### 【図2】

音声データ生成部の説明図である。

##### 【図3】

音声信号説明図である。

##### 【図4】

図4は集計／比較部の構成図である。

##### 【図5】

第2実施例の音声認識装置の構成図である。

##### 【図6】

従来の音声認識装置のブロック構成図である。

##### 【図7】

音声区間、非音声区間の説明図である。

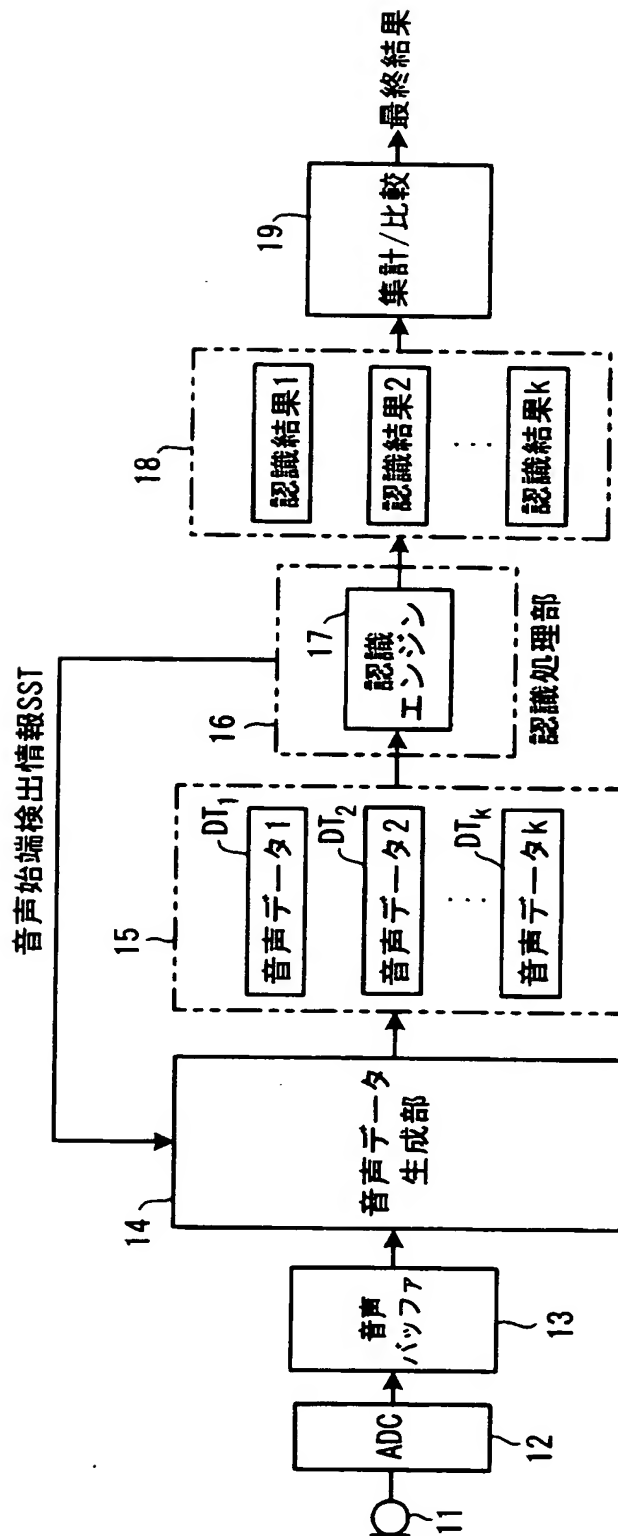
【符号の説明】

- 11 音声入力用マイクロホン
- 12 ADコンバータ
- 13 音声バッファ
- 14 音声データ生成部
- 15 音声データ記憶部
- 17 音声認識エンジン
- 18 認識結果保存部
- 19 集計／比較部

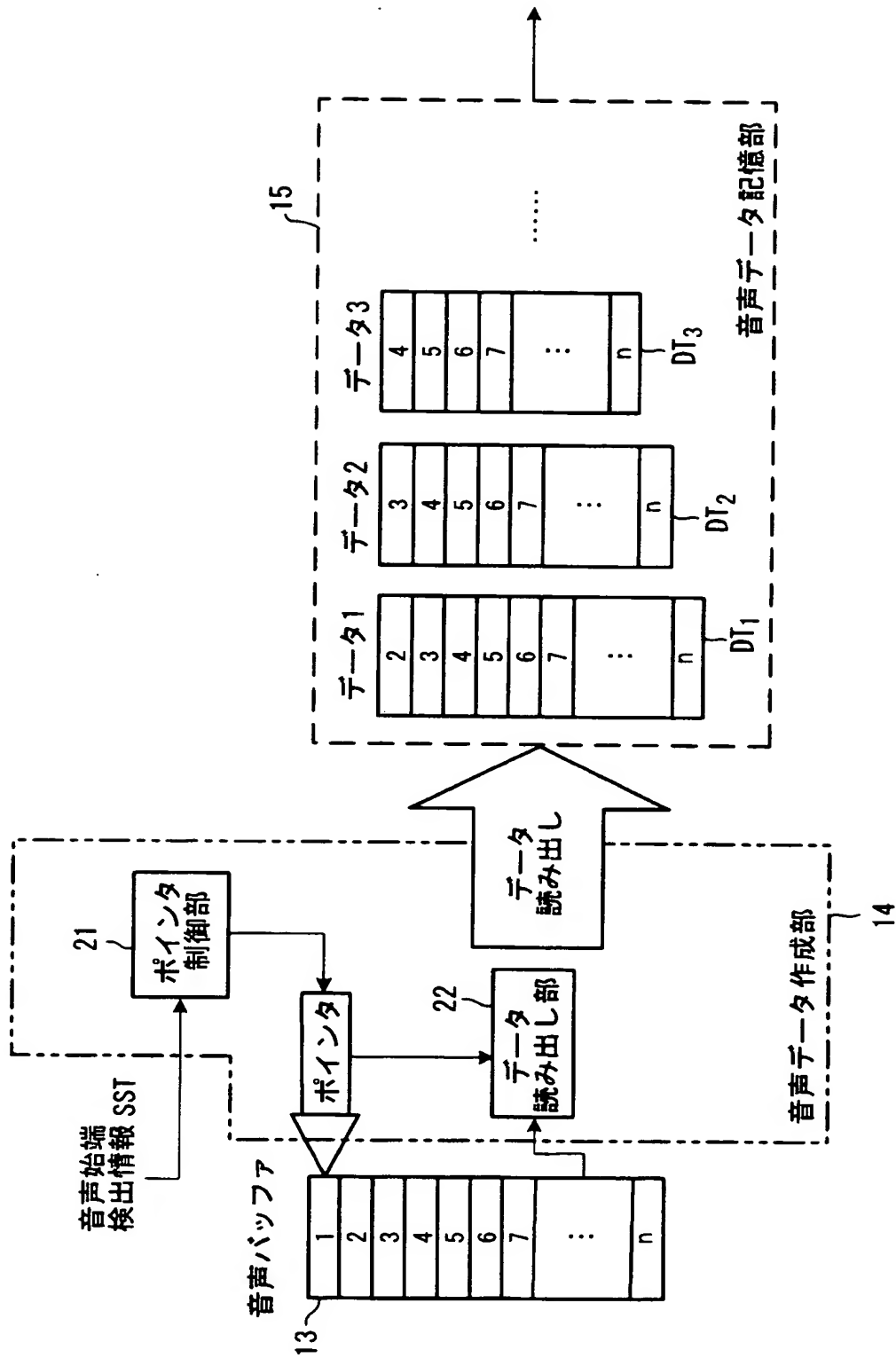
【書類名】

図面

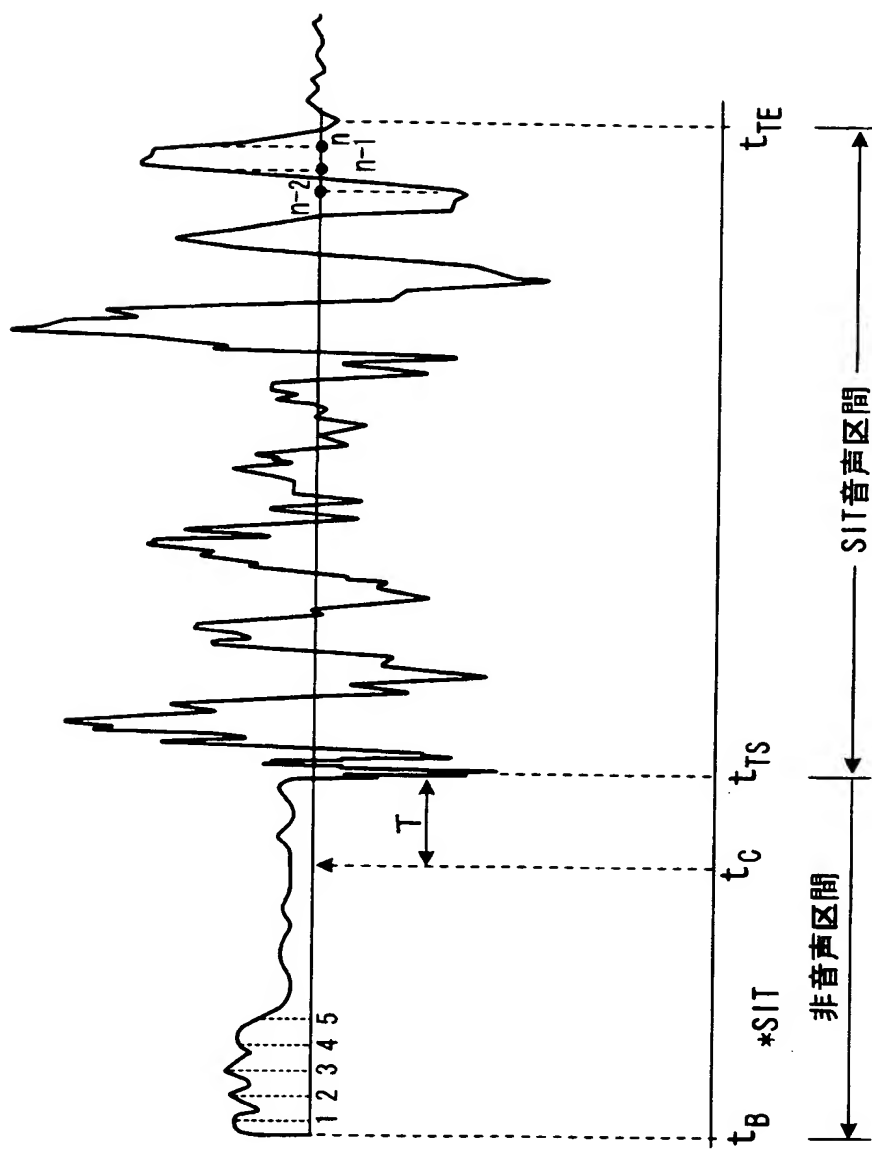
【図 1】



【図 2】

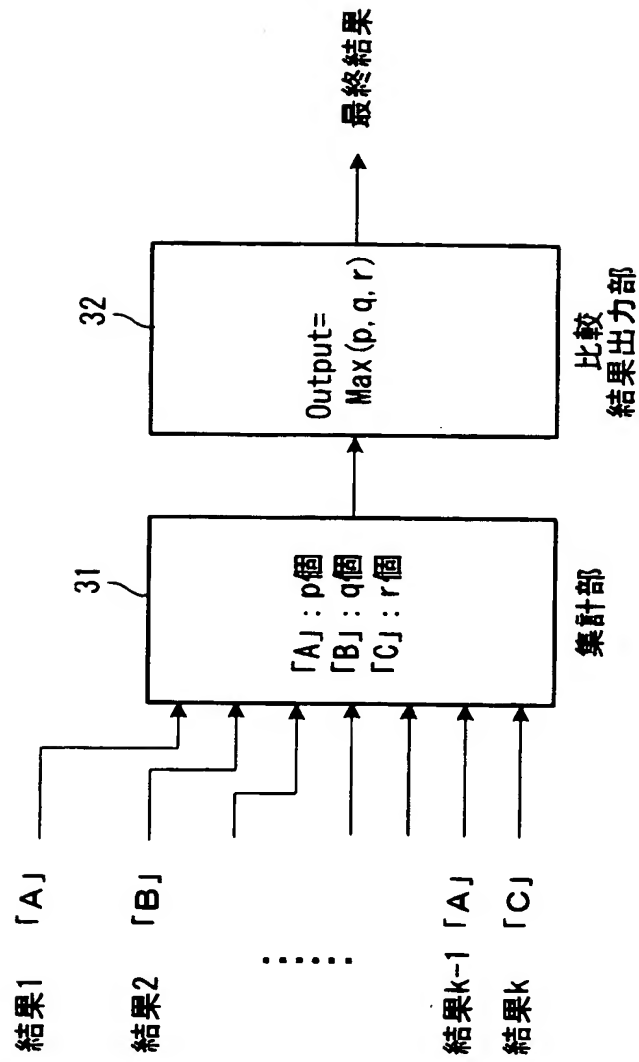


【図 3】

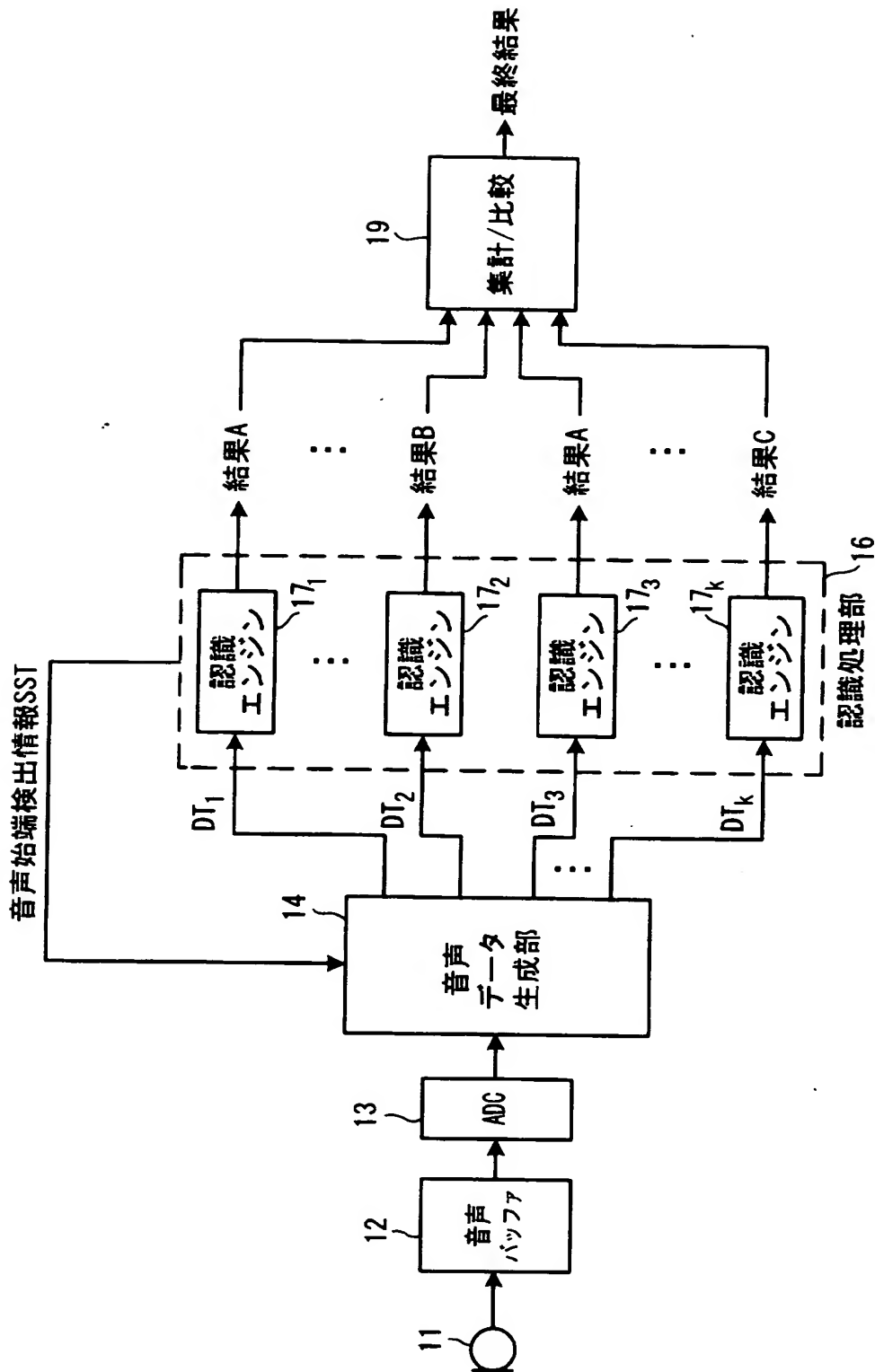




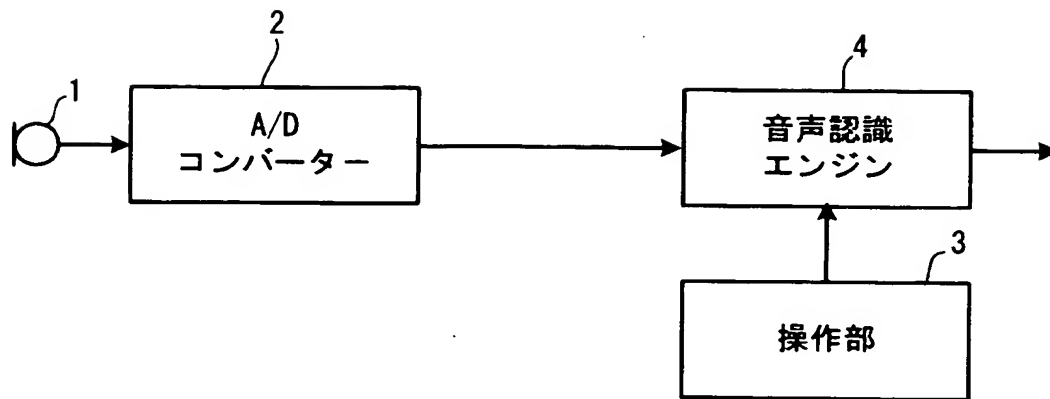
【図 4】



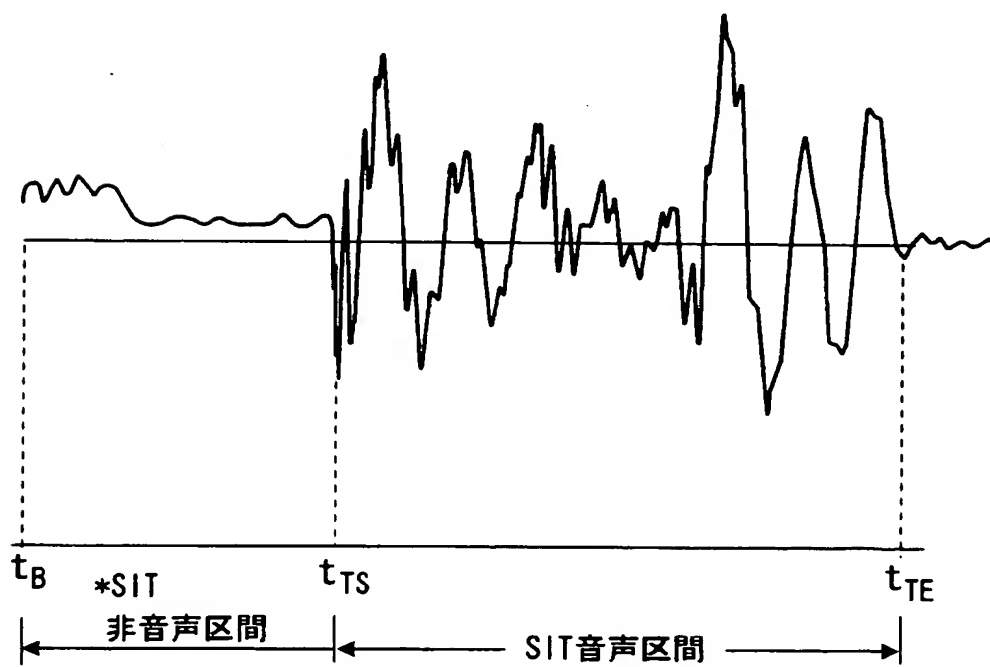
【図 5】



【図 6】



【図 7】



【書類名】 要約書

【要約】

【課題】 認識エンジンを変えずに音声認識性能を向上する。

【解決手段】 音声データ生成部14は音声認識対象である音声データより非音声区間の始点位置が異なる複数の音声データを生成し、認識エンジン17は各音声データを用いて音声認識を行い、集計／比較部19は得られた複数の認識結果のなかで、最も多い認識結果を出力する。

【選択図】 図 1

特願 2 0 0 2 - 3 6 0 3 3 6

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 1 0 1 7 3 2 ]

1. 変更年月日

1 9 9 0 年 8 月 2 7 日

[変更理由]

新規登録

住 所

東京都品川区西五反田 1 丁目 1 番 8 号

氏 名

アルパイン株式会社